OSU Working Papers in Linguistics 44 (191-223)

The influence of syntax on prosodic structure in Japanese*

Jennifer J. Venditti
venditti@ling.ohio-state.edu

Abstract: This paper examines the relationship between the syntactic and prosodic structures of utterances which are structurally ambiguous. Two experiments were conducted involving ambiguous noun phrases (right- versus left-branching) and relative clause constructions containing an adjunct with ambiguous scope of modification. Results of careful examination of the F0 contours and downstepping patterns reveal that inter- and intra-speaker variability as well as depth of syntactic embedding are important factors in determining the prosodic phrasing.

1 Introduction

Many acoustic studies in recent years have been concerned with the role suprasegmental features play in helping cue the syntactic structure of a sentence (e.g. Cooper et al., 1978; Klatt, 1975; Lehiste, 1973; Lehiste et al., 1976; Streeter, 1978; Terken & Collier, 1992; among many others). There is a wealth of literature concerning the disambiguation of syntactic ambiguity, discussing acoustic features such as duration, amplitude and fundamental frequency (F0) which provide a crutch to cueing differing interpretations of an otherwise ambiguous string of segments. Klatt (1975) found in English that increased duration can mark the ends of major syntactic units. Streeter (1978) showed that both duration and pitch contour, as well as amplitude, serve as salient cues to determining syntactic boundaries in ambiguous algebraic expressions. Lehiste (1973, 1976) found that duration, and F0 to a certain extent, plays a major role in disambiguation of many syntactically ambiguous sentences.

Japanese is no different from the Indo-European languages in this respect: there has been much work done on ambiguity which shows that acoustic features can indeed cue the syntactic structure (e.g. Uyeno et al., 1979, 1980, 1981; Azuma & Tsukuma, 1990, 1991; Venditti & Yamashita, in preparation). A study done by Uyeno and her colleagues (1980) on relative clause constructions in Tokyo Japanese showed a distinct F0 contour for each of the interpretations of an ambiguous utterance. An example sentence from their corpus is given in (1).

(1)

Ototoi	koronda	otona-ga	waratta
day before yesterday	fell	adult-NOM	laughed

A: 'The adult who fell the day before yesterday laughed.'
B: 'The adult who fell laughed the day before yesterday.'

*The work reported in this paper was supported in part by an Ohio State University Center for Cognitive Science Interdisciplinary Summer Fellowship and a Title VI Foreign Language Area Studies Fellowship. I would like to thank Mary Beckman, Beth Hume, Stefanie Jannedy, Keith Johnson, Sun-Ah Jun, Andreas Kathol, Mineharu Nakayama, and Frederick Parkinson for their helpful discussions, and Azusa Morii for her native speaker talents and endless cooperation.

From a perception test using synthesized F0 contours containing no pauses, it was shown that when the F0 peak on the adverb *ototoi* 'the day before yesterday' (Peak 1) is a great deal higher than the peak on the verb *koronda* 'fell' (Peak 2), interpretation A (in which the adverb is modifying the verb of the relative clause) is preferred. On the other hand, manipulation such that Peak 2 is equal to or higher that Peak 1 will yield a tendency toward interpretation B, in which the initial adverb modifies the verb of the matrix clause. This result has been replicated by Azuma and Tsukuma (1990, 1991) for the Tokyo dialect, and for the Kinki dialect as well.

The studies mentioned above all assume that variations in the suprasegmental characteristics of the speech signal are direct manifestations of differences in syntactic structure. They all assume that the syntax influences the phonetic representation in a straightforward fashion, without any mediating levels of structure. However, work in the last decade on the relation between the phonetic representation and the surface syntax has suggested there is indeed an intermediate level: the prosodic structure (e.g. Selkirk, 1984, 1986; Nespor & Vogel, 1986; Kubozono, 1988; among many others). The proposal of such a structure was motivated by observations that the domains of various phonological rules are not isomorphic to the syntactic structure, but in fact only correspond to the syntax in an indirect manner. This has in turn led to a handful of syntax-prosody mapping algorithms, as will be briefly discussed in §5. It is generally accepted now that there is a prosodic structure which is made up of hierarchically organized levels of phrasing, each of which corresponds to the domains of phonological phenomenon like downstep application in English or Japanese, or postlexical gemination in Italian. Prosodic constituents at each level of the hierarchy are independently motivated according to which phonological rules apply within their domain.

The existence of an intermediate prosodic structure which is only indirectly related to the syntactic structure creates the need to reconsider the results of previous acoustic studies on disambiguation cited above. It is apparent that the syntactic structure in some way has an effect on the phonetic output of the utterance, but *how* it has this effect is not clear. Do the different interpretations of ambiguous sentences have distinct prosodic representations? If so, at what levels in the prosodic hierarchy are the two distinct? The present study attempts not only to show that there are indeed differences in the speech signals in ambiguous constructions, but also to examine whether these are different in terms of their prosodic representations, and if so, to determine which levels of the hierarchy are relevant for disambiguation in Japanese.

2 The Prosodic Hierarchy in Japanese

Before proceeding to description of the experiments and discussion of prosodic representations of ambiguous constructions, I will give a whirlwind tour of those parts of the prosodic hierarchy of Japanese which have, to a large extent, been agreed upon by many of those working with Japanese intonation (see Beckman & Pierrehumbert, 1986; Kubozono, 1988, 1989, 1992; Maekawa, 1991; McCawley, 1968; Poser, 1984; Pierrehumbert & Beckman, 1988; Selkirk and Tateishi, 1988, 1991; among others).

I will assume throughout this paper that the intonation pattern is intimately connected to the prosodic structure of an utterance, and that this structure is directly manifested by the surface realization of the fundamental frequency contour. Thus, by observing patterns in this contour, we are able to make claims about the prosodic organization of the utterance.

2.1 Lexically specified pitch accent

In Japanese, each lexical item is classified as accented or unaccented. An 'accented' word is one which has a bitonal H*+L (henceforth HL) pitch accent associated to some specified mora. It is generally accepted that a word can have a maximum of one accent associated to it (see Poser (1990) for argument), and the location of the accent is determined at the lexical level. An 'unaccented' word, on the other hand, does not have this pitch accent associated to it, and is characterized mainly by tones associated with the accentual phrase level of the prosodic hierarchy (see §2.2). Consider the following minimal pair:

(2) accented: ue'ru mono 'the ones that are starved'
HL

unaccented: ueru mono 'something to plant'

(Henceforth I will transcribe an accented word with the diacritic (') after the mora to which the accent is associated.) Thus, the HL bitonal accent which is characteristic of accented words is assigned in the lexicon, while other tones which characterize a word, be it accented or unaccented, are assigned at the accentual phrase level of the intonational structure.

2.2 The accentual phrase: Phrasal and boundary tones

The accentual phrase is a level of the prosodic hierarchy which is essentially the same entity as the 'minor phrase' discussed by other scholars. This level of prosodic phrasing is defined for Japanese as "the domain of a postlexical rule deleting all accents after the first in an accentual phrase, and, more important, it is the domain of two delimitative peripheral tones, the phrasal H and the boundary L%." (Pierrehumbert & Beckman, 1988:26) An accentual phrase is most commonly thought to consist of a word plus its postposition or case marker. However, it is quite possible for a sequence of more than one word to combine to form one accentual phrase delimited by the phrasal H and L% boundary tone, and optionally at most one lexically specified accent. Figure 1a shows an accentual phrase consisting of a two-word sequence with an accented lexical item *ue'ru* 'to starve'. The fundamental frequency contour (the manifestation of these tones) is characterized by an initial L% boundary tone inserted in absolute utterance-initial position, a H phrasal tone followed by a steep fall associated with the HL pitch accent, and finally by the L% boundary tone. Figure 1b shows the contour of a single accentual phrase consisting of a sequence with an unaccented lexical item *ueru* 'to plant'. Here, the fundamental frequency starts off low (utterance initial L%), rises to the H phrasal tone and gradually tapers off toward the final L% boundary tone. This H phrasal tone will be associated to the second mora (if it is a short syllable) of a word.

The grouping of words into accentual phrases is a complicated phenomenon which is beyond the scope of this paper. It is worth mentioning, however, that accented words tend to resist grouping with other accented words into a single accentual phrase, while unaccented words tend to group together more readily. This fact will become relevant below when the intonation contours are examined. (For a discussion of factors governing accentual phrase formation, see Kori (1992) for Japanese and Jun (1993) for a related discussion of Korean.)

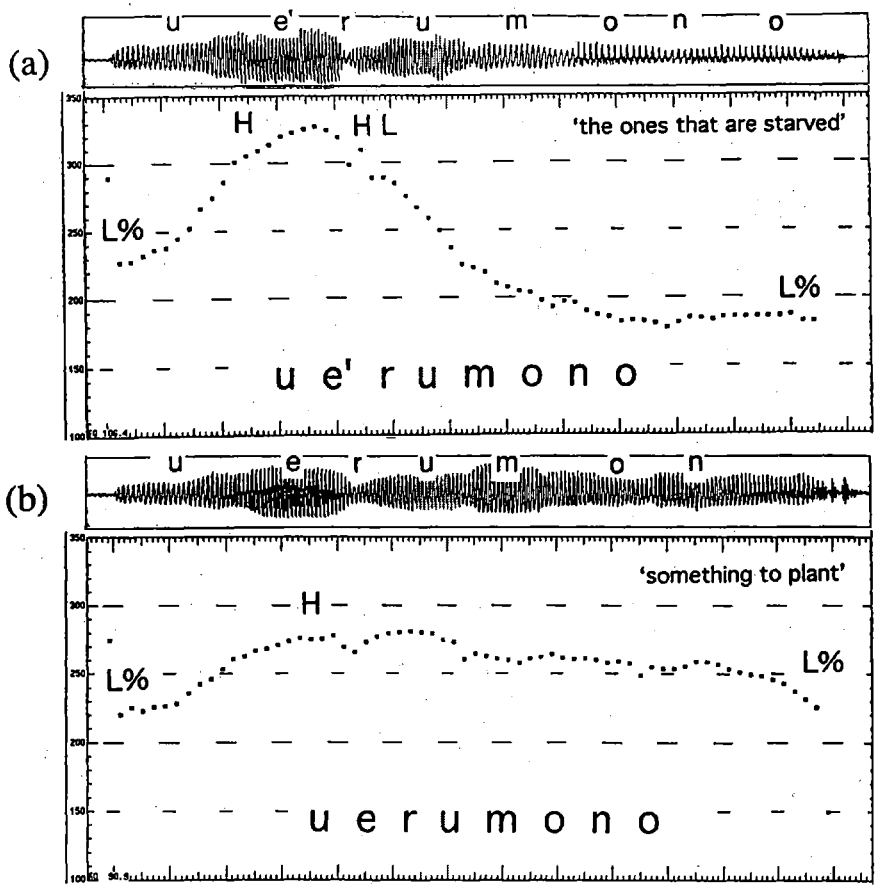


Figure 1. (a) An accentual phrase with an accented word *ue'ru mono* 'the ones that are starved'. (b) An accentual phrase with an unaccented word *ueru mono* 'something to plant'.

2.3 The major phrase: The domain of downstep

Much as in English and many African tone languages, pitch range in Japanese is manipulated by a process called downstep (or 'catathesis' as termed by Poser (1984) and Pierrehumbert & Beckman (1988)). In Japanese, downstep is the phonologically conditioned reduction in pitch range after the HL pitch accent. It has been suggested by Poser (1984) and confirmed by Pierrehumbert and Beckman (1988) and Kubozono (1988) that downstep applies iteratively within the bounds of a prosodically grouped set of accentual phrases. This larger prosodic constituent has been called the major phrase, and corresponds to Pierrehumbert and Beckman's 'intermediate phrase' which they compare to the analogous domain of downstep in English. Thus, in a string of accented accentual phrases that together form a major phrase, the pattern of the accentual phrase peaks will resemble a descending staircase. By definition, the contour of a string of unaccented accentual phrases

will not show the same pattern, since there are no accents to trigger the downstep process.

The level of the major phrase is extremely relevant to the present study, as the patterning of downstep between successive peaks will be examined in order to determine the proper prosodic phrasing of the utterances.

2.4 Contrasting theories of prosodic organization

In contrast to the widespread agreement on which levels of the prosodic hierarchy are relevant for Japanese, there is less uniformity of opinion about how the levels are organized with respect to each other. There are two main viewpoints: that held by Kubozono (1988, 1989, 1992) and that held by Beckman and Pierrehumbert (Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988).

Kubozono follows Ladd (1986) in assuming a recursive prosodic structure. That is, he explicitly rejects the Strict Layer Hypothesis (see Selkirk, 1984) by which prosodic units of type X^{n-1} are exhaustively grouped into units at the next higher level X^n, and instead proposes that prosodic constituents in Japanese are arranged in a binary branching hierarchical structure, which he gets by allowing embedding of prosodic constituents within constituents of the same type, shown in Figure 2a. This idea was originally proposed by Ladd for English to account for observed trends of 'declination within declination' — that is, downward trends of the fundamental frequency contour which seem to be embedded within larger downward trends. The idea was further developed by Ladd (e.g. 1988, 1990, 1993) into a metrical representation of prosodic structure and pitch register. In his model, the local prominences of pitch accents are represented as high and low terminal nodes of a metrical tree which can in turn be dominated by other high or low mother nodes to show prominence relationships among groups of accents. Such a representation attempts to account for relations among pitch registers by the combination of highs and lows and by the depth of embedding in the metrical tree. In such a model, there are necessarily a fixed number of pitch range values.

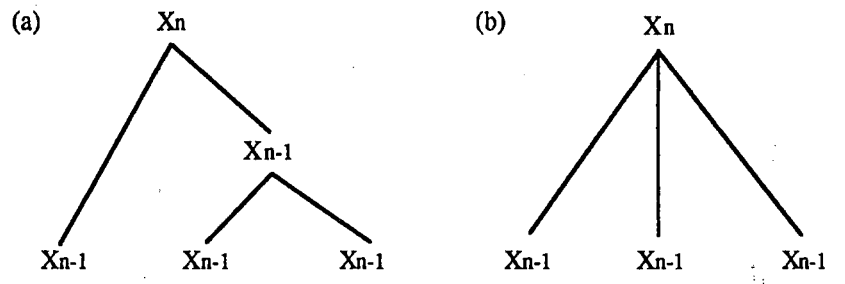


Figure 2. (a) Recursive versus (b) strictly layered hierarchical representations.

An alternate model of prosodic structure and pitch range relationships is described by Beckman and Pierrehumbert. They propose an n-ary branching hierarchical structure in which prosodic constituents group together according to the Strict Layer Hypothesis, as shown in Figure 2b. In this model, pitch range is a continuously variable phonetic specification chosen just once for any given major or 'intermediate' phrase, and at the beginning of a new major phrase, the pitch range is specified independently from that of the preceding phrase. Thus, the choice of pitch range for a given phrase is a paradigmatic one, reflecting the overall discourse structure or the pragmatic context. This forms a sharp contrast to the proposals of

Ladd and Kubozono, in which pitch range relationships are represented phonologically as a result of particular arrangements in the structure of their prosodic representation.

2.5 Differing accounts of ambiguous constructions

With these differences in mind, let us now turn to examples of syntactically ambiguous constructions.

Kubozono (1988, 1989, 1992) has studied extensively the relation between syntactic structure and fundamental frequency peak values for noun phrases with differing branching structures, such as the those given in (3). He also examined ambiguous strings such as in (4).

(3)	a.	[[ao'yama-ni	a'ru]	daigaku]	
		Aoyama-in	exist	university	'a university in Aoyama'
	b.	[ao'yama-no	[a'ru	daigaku]]	
		Aoyama-GEN	certain	university	'a certain university in Aoyama'
(4)	a.	[[o'okina	nooen-no]	o'onaa]	
		big	farm-GEN	owner	'an owner of a big farm'
	b.	[o'okina	[nooen-no	o'onaa]]	
		big	farm-GEN	owner	'a tall farm-owner'

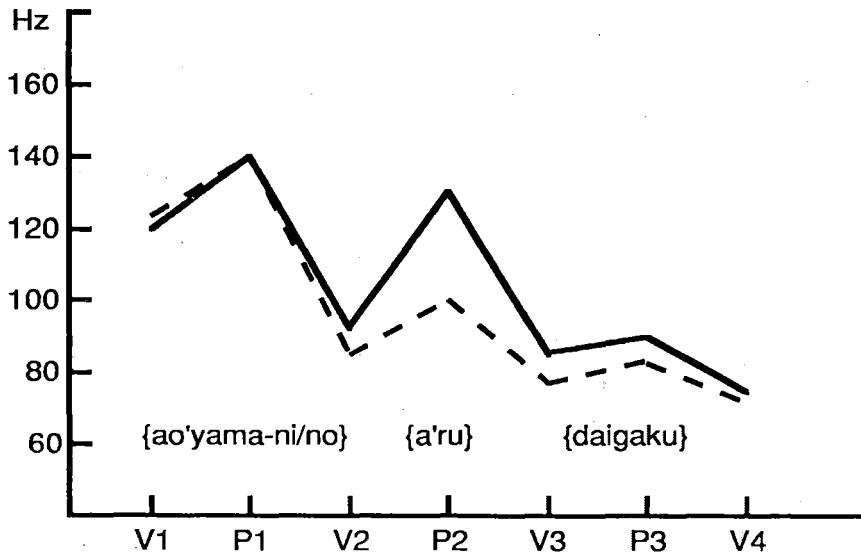


Figure 3. A schematization of the intonation contour of a left-branching (dotted line) versus right-branching (solid line). (Adapted from Figure 15.4 in Kubozono (1992)).

Kubozono's data show that, while in the left-branching structure ((3a) & (4a)) the F0 peak on the first word (Peak 1) is a great deal higher than that on the second

(Peak 2), the right-branching structure ((3b) & (4b)) yields a pattern in which the height of Peak 2 is about the same as that of Peak 1. This pattern of the relative heights between the first two peaks is similar to those observed in the studies by Uyeno et al. and Azuma and Tsukuma described above. Figure 3 is a schematization of Kubozono's mean peak (P) and valley (V) measurements for many tokens of the left-branching (LB) and right-branching (RB) structures in (3). Kubozono not only notes a distinct physical difference in the contours for the two branching structures, but uses these observations to motivate a prosodic representation of utterances such as these. He proposes two contrasting prosodic structures, shown in Figure 4a & b. By using a recursive structure in which minor phrases (mp) are embedded within minor phrases, Kubozono is essentially encoding the difference in syntactic branching directly into the phonological representation of these noun phrases. As I will describe below, he claims that it is necessary for the phonology to be able to access the syntactic structure in order to describe the differing peak height relationships which are observed (cf. Figure 3).

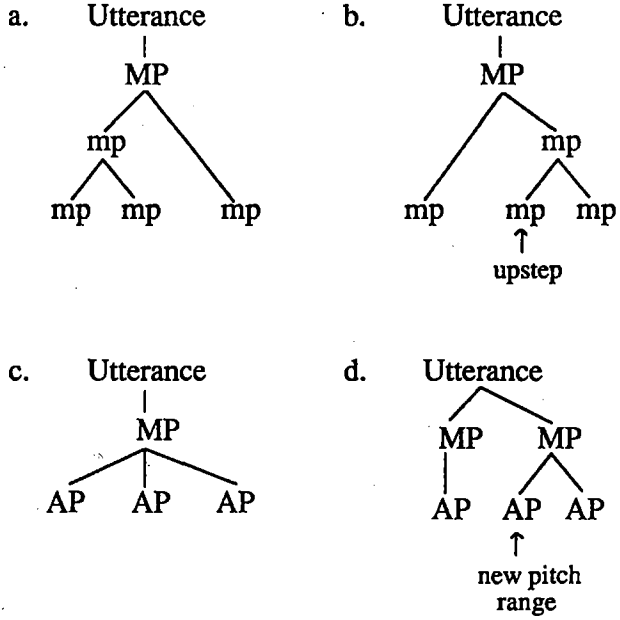


Figure 4. Prosodic structure of (a) left-branching and (b) right-branching noun phrases as proposed by Kubozono (Adapted from example (13) in Kubozono (1992)), and possible (c) left-branching and (d) right-branching structures according to the Beckman and Pierrehumbert model.

By comparing utterances with accented and unaccented initial words, Kubozono (1988, 1989, 1992) argues that downstep, the reduction in pitch range triggered by an accent, is indeed occurring between Peak 1 and Peak 2 (minor phrases 1 and 2) in both of the branching structures shown in Figure 3. The application of this phenomenon will cause the height of Peak 2 to be lower relative to Peak 1 in both cases. While this is an adequate description of the observed downward trend in the LB structure, it is clear that in the RB structure the second peak is higher than it would have been if only downstep had applied. Instead of proposing two types of

downstep, Kubozono introduces the mechanism of 'metrical boost' — a local boost in fundamental frequency which applies to the leftmost constituent of a right-branch. He treats a right-branching structure as 'marked' in a predominantly left-branching language like Japanese. Thus, this phonetic upstep mechanism (see arrow in Figure 4b) which he proposes is only indirectly sensitive to syntactic branching via the prosody. To recapitulate then, Kubozono claims that downstep is occurring between Peaks 1 and 2 in both branching structures, but there is an additional upstep mechanism of 'metrical boost' applied to the peak in the RB structure which corresponds to the leftmost minor phrase of a right-branch.

The alternative model of prosodic organization proposed by Beckman and Pierrehumbert might provide a different explanation for the pitch range scaling of Peaks 1 and 2 in the noun phrases discussed above. Though they have not made claims about these particular constructions in Japanese, it is reasonable to presume that the different fundamental frequency patterns of the left-branching and right-branching structures shown in Figure 3 might be accounted for by a difference in the major (or 'intermediate') phrasing for the two. In the LB structure, such as that in (3a) & (4a), the utterance might be composed of one major phrase, within which downstep chains to resemble a staircase like pattern, whereas the RB structure, as in (3b) & (4b), might have a new major phrase starting before Peak 2 so that the application of downstep is blocked. The pitch range selected for the new phrase would reflect the pragmatic relationship between the two phrases. Prosodic structures of the two branching structures under this account are also schematized in Figure 4 (c & d).

2.6 Objectives of the present study

Although it is clear by all previous descriptions of ambiguous utterances in Japanese that differing syntactic structures can indeed be disambiguated by means of the intonation, there is not total agreement as to *how* this is done, as was seen above. The description of the prosody of left-branching structures is fairly straightforward in all accounts, but the prosodic representation of right-branching structures has yet to be agreed upon. One possibility is that RB structures such as (3b) & (4b) are comprised of one major phrase with a syntactically sensitive metrical boost applying to the leftmost constituent of a right branch, while another possibility is that there are two major phrases with a pitch range reset at the boundary. Both of these accounts can yield a similar contour, such as that shown by the solid line in Figure 3, however, each account makes different assumptions about the application of downstep. Kubozono's account assumes that downstep occurs between the first and second peaks in both LB and RB structures. The model of Beckman and Pierrehumbert, in contrast, would say that downstep between Peak 1 and Peak 2 will be blocked in the RB structure. Whether there is downstep or not is an empirical issue which deserves to be examined more closely. In addition, since syntactic branching structure is deeply woven into Kubozono's recursive prosodic representation (which in turn determines factors like downstep and metrical boost), a major prediction of his account is that there is only one possible way for speakers to produce RB constructions. Beckman and Pierrehumbert's account does not necessarily predict the lack of consistency, but it is less constrained, in that it can potentially allow for more variation in relative peak heights depending on pragmatic influences. Therefore, it will be of interest to know whether speakers are consistent in their productions of the RB utterances.

Also, Kubozono's discussion is limited to the representation of simple branching structures such as those in the noun phrases. Since similar phenomenon have been noted in ambiguous strings involving more complex constructions such

as relative clauses as well (cf. (1)), we would want a model which would account for these cases in a similar way. Beckman and Pierrehumbert might describe the more complex constructions in a similar way, making 'right-branching' relative clauses like that in (1b) analogous to the right-branching noun phrases, which contain two major phrases. Kubozono, on the other hand, suggests that metrical boost is an n-ary process which will apply multiply according to the depth of the right-branching structure, thus accounting for the larger pitch rise in more complex structures with deeper embedding. In order to give a concrete general account of the prosodic representation of these structures, however, the patterning of downstep must be examined in detail for both noun phrases as well as more complex structures.

This paper discusses the results of two experiments designed to compare these two models of prosodic structure. Both ambiguous noun phrases and relative clause constructions involving an adverb (or locative adjunct) with ambiguous scope of modification were elicited from native speakers of the Tokyo dialect. Fundamental frequency contours were extracted and peak measurements were made in order to examine the behavior of downstep in each construction. Perception experiments for each corpus were also done to confirm the production results. Results show that for the RB noun phrases, in which the syntactic boundary of the right-branching element is not so deep, two of the three speakers exhibited a pattern of downstepping between the first two peaks, as predicted by Kubozono's model. The third speaker showed a downstep reset on the second peak, in accordance with the model proposed by Beckman and Pierrehumbert. This lack of consistency may be due to different speaker strategies for disambiguation. In contrast, for the relative clause constructions, in the right-branching structure (cf. (1b)) whose leftmost edge is deeply embedded, the results for all speakers favored the analysis whereby downstepping is blocked on Peak 2 and the pitch range is reset in the new major phrase. Also, a pause was present in each right-branching structure between Peaks 1 and 2 to aid in the disambiguation. These results suggest that individual speaker strategies or inter-speaker variability as well as depth of embedding are important considerations that must be addressed when proposing a general account of the prosodic structure of Japanese and its interaction with the syntax.

3 The experiments

3.1 Experiment 1 — Noun phrases

3.1.1 Production

Rationale
This experiment was conducted in an attempt to replicate Kubozono's (1988) findings that downstep occurs between Peaks 1 and 2 in both left-branching and right-branching noun phrases. In contrast, Beckman and Pierrehumbert's model suggests that, while downstep will apply between these two peaks in the LB structure, it will be blocked in the RB structure.

Methods
Three native speakers of Tokyo Japanese were recorded in a double-walled sound booth at the OSU Linguistics Laboratory. The task in this experiment was to describe various pictures which were mounted on a wall in the booth. The experimenter would point to a picture while saying the prompt *kore-wa nan desu ka?* 'What is this?', and the speaker would then respond using the appropriate noun

phrase in the carrier sentence *sore-wa .... desu* 'That is ....'. The pictures depicted differing interpretations of ambiguous segmental strings. The noun phrases used are given in (5).

(5)

kimi'dorino	hima'wari-no	moyoo	aiirono	hima'wari-no	moyoo
green	sunflower-GEN	pattern	indigo	sunflower-GEN	pattern
kimi'dorino	kanariya-no	moyoo	aiirono	kanariya-no	moyoo
green	canary-GEN	pattern	indigo	canary-GEN	pattern

Each of the noun phrases was controlled for phonological length (number of morae) of its components as well as vowel height on the relevant morae. This allowed for a direct comparison of peak height. All combinations of accentedness for the first and second words (i.e. +A+A, +A-A, -A+A, -A-A) were included. Each of the four noun phrases is ambiguous in that it has two possible interpretations depending on the branching structure. The left-branching structure would be interpreted as 'the pattern of green / indigo sunflowers', and the right-branching would be 'the green / indigo pattern of sunflowers'. In a preparation session, the speakers were given a concrete context in which these noun phrases might appear, and were informed of the ambiguity involved. They were then asked to describe unambiguously the scene depicted in the picture (there were two pictures per single noun phrase). Ten tokens were elicited for each interpretation for all of the noun phrases for all speakers, resulting in a total of 240 utterances (3 speakers x 2 branching x 2 word1 accentuations x 2 word2 accentuations x 10 tokens). The utterances were elicited in random order.

The utterances were then digitized at 10KHz (12 bit resolution) and the fundamental frequency contour extracted for each token using an autocorrelation-based F0 tracker. The pause duration between offset and onset of voicing of words 1 and 2 was measured, as well as the fundamental frequency value for each peak in the utterance. For utterances in which the first and second words were accented, measurement of the peaks was straightforward. However, in cases where either the first or second word was unaccented, dephrasing often occurred (i.e. the two words were produced in a single accentual phrase), making it impossible to make a 'peak' measurement. Typical contours of the left-branching noun phrases are shown in Figure 5. It is clear from the example contours that there is a tendency for accented words to form their own accentual phrase separate from adjacent words (cf. (a)), while unaccented words tend to be dephrased together with surrounding words (thus not having distinguishable peaks) (cf.(b-d)). Therefore, it becomes impossible to check for downstepping between two peaks, such as Peaks 1 and 2, if they are phrased together. For this reason, the results presented below will only address the *hima'wari* type tokens, in which word 2 is accented ((a) & (c)). This assures that there will be a distinguishable Peak 2 whose height can be measured (see discussion of downstep below).

Results and Discussion

In order to examine the predictions of Kubozono's theory on the one hand, and those of Beckman and Pierrehumbert on the other, it is necessary to look at the downstep patterning in both the left-branching and right-branching interpretations of an ambiguous noun phrase. Example contours of these structures are given in Figure 6.

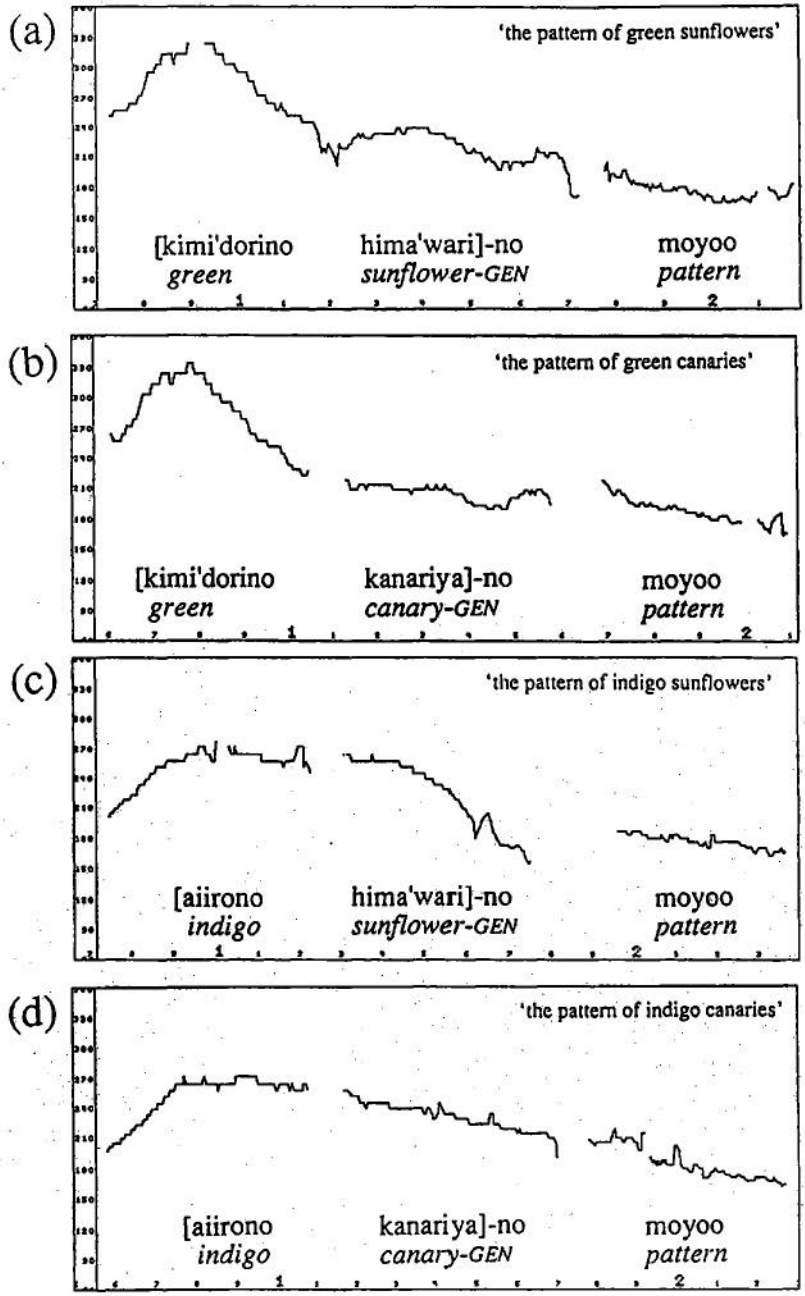


Figure 5. Example contours of left-branching noun phrases. Accentuation of first and second words: (a) +A+A, (b) +A-A, (c) -A+A, (d) -A-A.

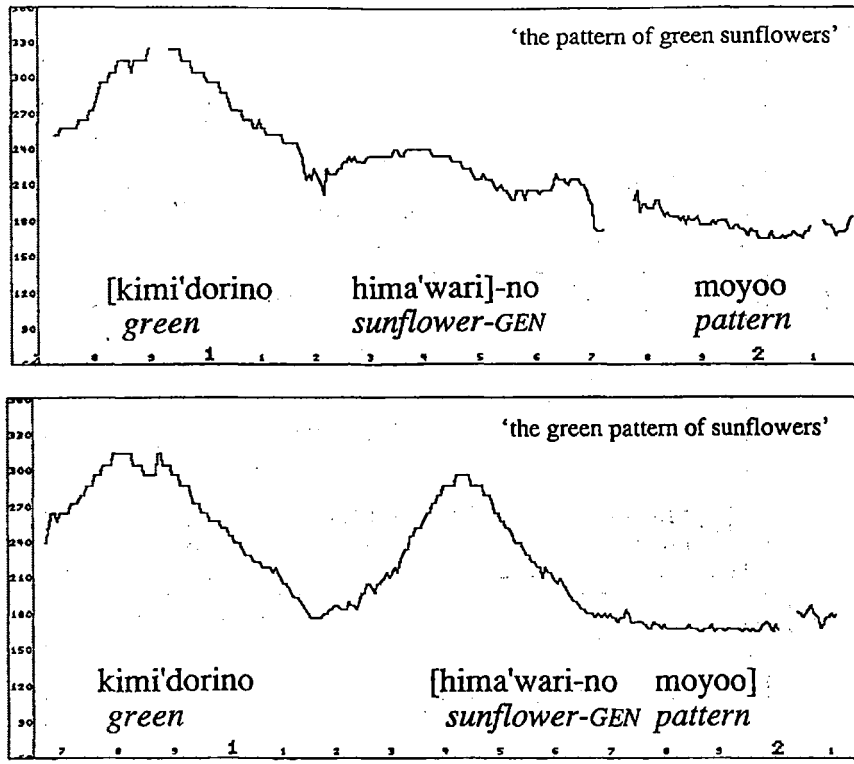


Figure 6. Typical fundamental frequency contours of left-branching (top) and right-branching (bottom) noun phrases with accentuation +A+A-A.

The LB structure in Figure 6a shows a downstepped pattern of a staircase-like descent. The RB structure, in contrast, does not show this pattern, but rather, the peak of the second element *hima'wari* is about as high as the peak on *kimi'dorino* (Peak 1). It is an empirical question then, whether downstep has applied to this peak on *hima'wari* in this structure and then been reversed by the application of metrical boost, or whether there is a major phrase boundary between Peaks 1 and 2 which blocks the application of downstep and causes a subsequent reset in pitch range. In order to test for the occurrence of downstep, it is not possible to look at only the degree of fall between two peaks in a given contour. Rather, since downstep is defined as the reduction in pitch range triggered by an accent, it is necessary to compare the heights of the target peaks (in this case *hima'wari*) when an accented versus unaccented word precedes them (here *kimi'dorino* vs. *aiirono*). If indeed downstep occurs between two peaks, such as between Peak 1 and Peak 2 here, then the prediction is that the height of Peak 2 would be significantly lower when preceded by an accented item than when preceded by an unaccented item. If, on the other hand, there is no downstep occurring between the two peaks, we would expect to see no significant difference in the height of Peak 2 as a function of the accentedness of the preceding word.

Since, as mentioned above, unaccented words tend to phrase together with adjacent words, it becomes impossible to look at the effects of downstep in LB structures (in which dephrasing occurred frequently) — that is, while the accented *kimi'dorino* forms its own accentual phrase whose peak height is readily measurable (cf. Figure 4a), the unaccented *aiirono* phrases together with *hima'wari* (cf. Figure 4c) and thus a 'peak' is not readily measurable. Therefore, I will focus on RB structures only in this examination of downstep between Peaks 1 and 2. Recall that the two contrasting theories of Japanese prosodic organization offer essentially the same account of the LB structures, but differ in their accounts of RB structures.

The graph in Figure 7 shows frequency values of the second peak height (*hima'wari*) when following an accented word *kimi'dorino* versus an unaccented word *aiirono* for one speaker (AM).

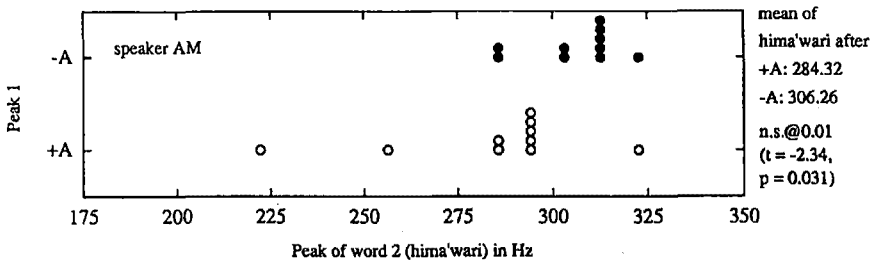


Figure 7. Fundamental frequency values of Peak 2 in the right-branching noun phrase [kimi'dorino / aiirono [hima'wari-no moyoo]] 'the green / indigo pattern of sunflowers'. Full circles indicate word 1 is -A, hollow circles indicate word 1 is +A. Speaker AM.

This graph shows that, with the exception of two outliers (at 225 Hz & 260 Hz) which are clearly downstepped (Peak 2 is quite low), the majority of the tokens following the accented Peak 1 are not substantially lower than those following unaccented Peak 1. Although the average values differ by about 20 Hz, the results of a t-test analysis indicate that the means are not significantly different ($t = -2.34$, $p > 0.01$).[1] The two outliers may be taken to be of a different population from the rest of the tokens, in which case a t-test using all tokens is rendered inappropriate. However, even a comparison of the values excluding the two outliers shows that the samples are not significantly different ($t = -1.89$, $p > 0.01$). These results indicate that, for speaker AM, there is no downstep occurring between Peaks 1 and 2 in the majority of the cases, disputing Kubozono's claim that it actually does apply in such constructions. This absence of downstep would be interpreted within a framework like that proposed by Beckman and Pierrehumbert as an indication of the presence of a major phrase boundary between these two peaks. Therefore, it appears that for the majority of utterances of the RB structure for this speaker, she in fact produced two major phrases, across which downstep is blocked (cf. Figure 4d).

However, for the two other Tokyo speakers used in this study, the patterning of downstep resembles the findings of Kubozono. Figure 8 shows the same comparison of the height of Peak 2 when following an accented versus unaccented word, as shown for speaker AM above.

[1]The 0.01 level of significance was used in all of the t-tests in this study.

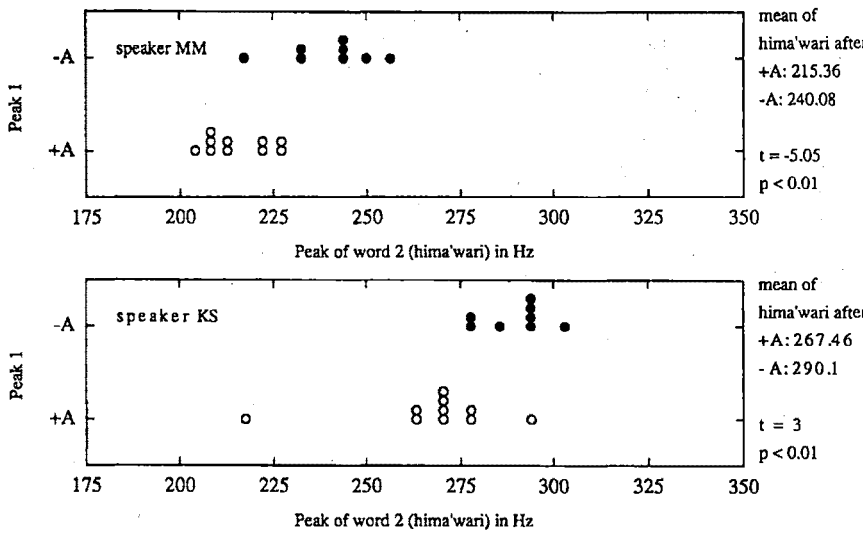


Figure 8. Fundamental frequency values of Peak 2 in the right-branching noun phrase [kimi'dorino / aiirono [hima'wari-no moyoo]] 'the green / indigo pattern of sunflowers'. Full circles indicate word 1 is -A, hollow circles indicate word 1 is +A. Speakers MM and KS.

In contrast to speaker AM, both speakers MM and KS showed a significant downstep relationship between Peaks 1 and 2 even in this RB structure. As can be seen from Figure 8, the height of Peak 2 when following an accented word is significantly lower (MM: $t = -5.05$, $p < 0.01$; KS: $t = 3.0$, $p < 0.01$) than when it is following an unaccented word. This is true for the data as a whole — they include no obvious outliers which we could say are phrased differently. This pattern of downstepping between Peaks 1 and 2 replicates the findings by Kubozono (1988, 1989, 1992), and supports his claim that the only difference between LB and RB structures is the presence or absence of metrical boost on the already downstepped leftmost constituent of a right branch.

Measurement of pause duration shows that, while all tokens of the left-branching structure were uttered with no pause between Peaks 1 and 2, the right-branching structures were equally divided as to whether there was a pause at this location or not. This indicates that the presence or absence of a pause alone cannot disambiguate the structures, but that indeed the contribution of the F0 contour plays a major role. This result supports Azuma and Tsukuma's (1990, 1991) findings that F0 is a more salient factor in disambiguation. Further discussion of the role of pauses in the present data will be addressed in the perception section 3.1.2.

This experiment involving ambiguous noun phrases was conducted in an attempt to replicate Kubozono's (1988) findings that downstep does indeed occur between the first and second peaks in right-branching constructions, a discovery which led him to propose the syntactically induced upstep mechanism of metrical boost. These results were replicated for two of the three Tokyo speakers used in the experiment, while the third speaker showed a behavior closer to that described by the Beckman and Pierrehumbert model. This suggests that, in the disambiguation of noun phrases, individual speakers may make use of different strategies,

involving variant prosodic phrasings or optional application of metrical boost. This variation suggests that it may not be appropriate to explain the 'boost' in peak height on the second peak in RB structures as something that is necessarily tied to the syntactic configuration (as is the case with metrical boost), but rather it can be thought of as an increased local pitch prominence and/or phrase break used by speakers to signal the disjuncture between the two adjacent words.

3.1.2 Perception

Rationale

In accordance with the proposals set forth by Uyeno et al. (1980) and Azuma and Tsukuma (1990, 1991), in which the height of the second peak compared to the first influenced the listeners' interpretation of the ambiguous sentence, it was predicted that in this perception experiment as well, there would be a similar correlation between the difference in height between Peaks 1 and 2 and listener judgment. Specifically, it was hypothesized that a large positive difference (Peak 1 is substantially higher than Peak 2) would cue the A interpretation (LB), while a negative difference or no difference (Peak 2 higher than or equal to Peak 1) would cue the B interpretation (RB). If this holds true, we should observe a positive linear correlation between the listener choice and the difference in height of the peaks.

Methods

Twelve native listeners of Tokyo Japanese participated in this experiment. Five tokens of each of interpretations A (left-branching) and B (right-branching) produced by each speaker were selected randomly from the *kimi'dorino hima'wari* and *kimi'dorino kanariya* types. This gave a total of 60 utterances (3 speakers x 2 branching x 5 tokens x 2 word2 accentuations). Each token was presented to the listener twice (randomly), making a total of 120 stimuli. The stimuli were transferred from digitized form onto an audio tape, and were presented in blocks of 15 utterances each, with the type of the token being constant within a block. The stimuli within each block were randomized, and played at 5 second intervals. Each occurrence of a token was heard only once. The subjects listened to the audio stimuli over headphones in a double-walled sound booth, with the relevant visual cues used in the production experiment (labeled A and B) mounted on a wall directly in front of them. Subjects were asked to listen to each utterance and judge which picture it was intended to describe on a five point scale from 'definitely interpretation A' through 'I don't know' to 'definitely interpretation B'. These choices were explained to the listeners before the start of the experiment by an instruction sheet written in Japanese. The pictures corresponding to A and B were reversed after half of the listeners had taken the test in order to avoid response bias due to order of responses. The participants were informed of the ambiguity of the sentences in a practice session beforehand and were asked to think of how they themselves might describe each picture unambiguously. No verbal prompt was given. There was a sample block prior to the actual test which was repeated as many times as needed in order for the listeners to feel comfortable with the task. At the close of the session, each participant was asked what cues they listened for in attempting to distinguish interpretation A from interpretation B. Listener judgments ranged from 54% (near chance level) to 95% correct.

Results

Figure 9 shows the listener choice (sum of choices for all listeners), ranging from 100% A judgments to 100% B judgments, plotted against the difference in ERB

(equivalent rectangular bandwidth) between Peaks 1 and 2 in each token. The ERB psychoacoustic measure was chosen since it has been shown to best reflect the scaling of pitch prominences in the perception of intonation (Hermes & van Gestel, 1991; Moore & Glasberg, 1983).[2] Qualitatively similar results can be found when substituting the difference in Hz or semitones for this measure.

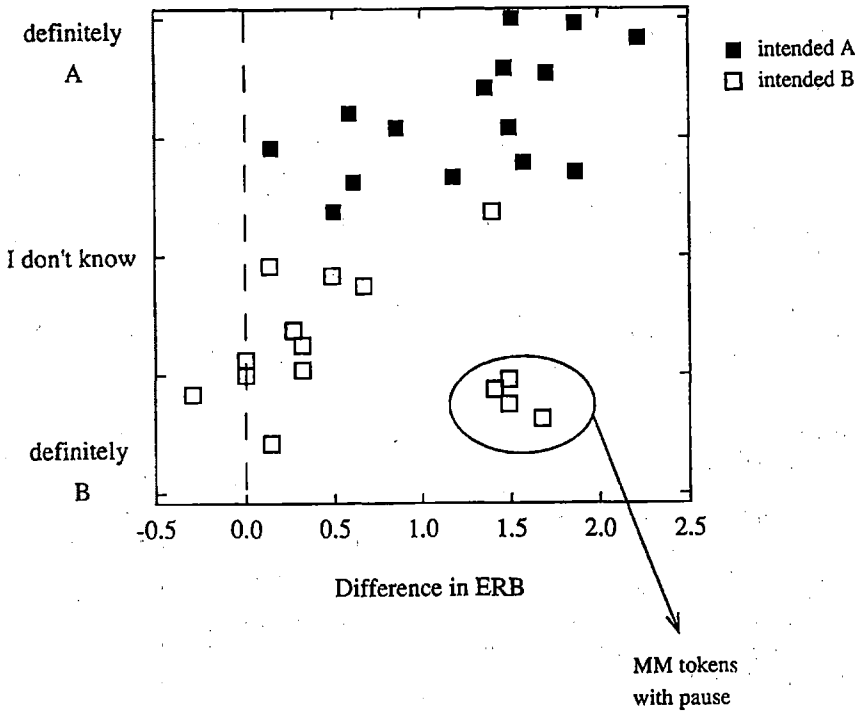


Figure 9. Listener choice from 'definitely A' to 'definitely B' plotted as a function of difference in ERB between Peaks 1 and 2 for [kimi'dorino hima'wari-no moyoo] noun phrases.

The data points in this graph show a gradual transition from 100% 'definitely A' judgments in the upper right-hand corner (greatest difference in ERB) to 100% 'definitely B' judgments in the lower left-hand corner (negative or no difference in ERB). These results support the hypothesis that the difference in the height of the peaks is directly correlated with the listener choice. The only apparent exceptions to this overall trend are the four outliers of the intended B type clustered around 1.5 ERB. Closer examination of these outliers shows an interesting effect concerning the role of the pause in judging these utterances. Each of these outlying tokens was produced by speaker MM, who made Peak 2 subordinate to Peak 1 (large positive difference in ERB) even in the right-branching constructions. The other speakers did not employ this strategy, though speaker MM seemed to use it often. The graph in Figure 9 shows that these outliers were judged by the native listeners as close to

[2]The equation used to calculate ERB-rate (equivalent rectangular bandwidth) is: $E = 11.17 \ln |(f+0.312)/(f+14.675)| + 43.0$ where f is in kHz (Moore and Glasberg, 1983).

the B interpretation in spite of this large peak difference, while the one token (intended B interpretation) which lies directly above them was judged near ambiguous or closer to the A interpretation. Examination of these specific tokens shows that, while the four outliers near the B end have a pause between Peaks 1 and 2, the one token closer to the A end does not include a pause.

This finding sheds light on the interaction between the relative peak height and the pause in cueing the structure of an ambiguous sentence. It seems that while a large difference in height between the two peaks does indeed play a major role in cueing the meaning, the pause has a nontrivial role also.

Figure 10 shows the relationship between the length of the pause between Peaks 1 and 2 and the listener judgment.

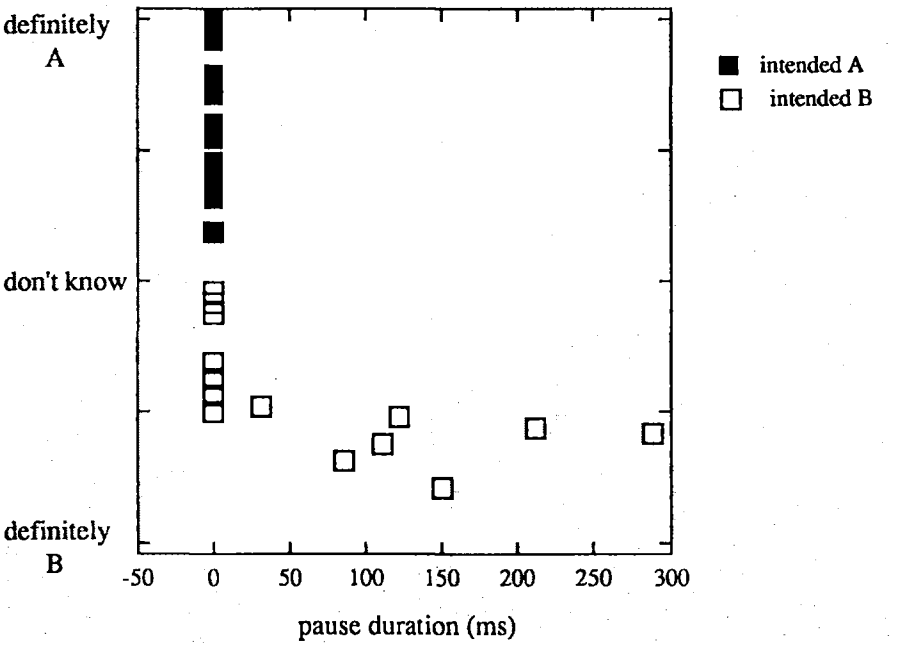


Figure 10. Listener choice from 'definitely A' to 'definitely B' plotted as a function of length of pause between Peaks 1 and 2.

This graph illustrates the fact that there is a regular relationship between the existence of a pause and the branching structure perceived by native listeners. It shows that, while all of the intended A tokens had no pause, the intended B tokens are split as to whether there was a pause or not. Within the intended B squares, those containing a pause between Peaks 1 and 2 were judged as being closer to interpretation B than those without a pause. This shows that the pause is a cue to the intended structure of an utterance. However, it is clear that the pause is not the only cue. Otherwise, we would expect to see all of the points for interpretation A overlaying each other, with no variation in the listener judgment, and all intended B tokens without a pause to be misperceived as A. As we can see from the graph, this is not the case — the tokens of intended interpretation A range from 'definitely A' to 'I don't know', and even with no pause no B tokens were judged to be very

A-like by the majority of listeners. This indicates that there are other factors beside the pause which are influencing judgments; namely, the difference in peak heights shown in Figure 9. In the short interview after the experiment, listeners were asked which cues they listened for in differentiating the two interpretations. The following is a summary of their reactions:

(6)	attended to pause (for B) only:	7 listeners
	attended to peak height difference only:	1 listener
	attended to both:	3 listeners
	neither:	1 listener

This suggests that the pause was the most salient cue to disambiguation. However, these impressions may not be too reliable, since there were many tokens which did not have a pause but still were correctly judged as type B tokens. It may be that the listeners were not conscious of all the cues involved, and just named the first which came to mind or the ones that they could most easily describe in words.[3]

In conclusion, results of this perception study show that while the difference in height of Peaks 1 and 2 is strongly correlated to the listener judgment, the presence or absence of a pause also plays a non-trivial part in disambiguation. This supports previous claims (e.g. Lehiste et al, 1976) that native listeners can take into consideration multiple prosodic factors in the perception of differing syntactic structures.

3.2 Experiment 2 — Relative clause constructions

3.2.1 Production

Rationale

In order to have a general account of the syntax-prosody relation for differing branching structures, it is necessary to examine structures which involve more complex branching than just LB and RB noun phrases. This experiment was designed to compare the accounts of the two alternate theories of Japanese prosodic organization for complex structures such as relative clause constructions with ambiguous scope of modification of temporal or locative adjuncts. Again, examination of downstepping patterns will be relevant for the assessment of the two accounts.

Methods

The three native speakers of Tokyo Japanese who participated in experiment 1 participated in this experiment as well. The task in this case was to read aloud a randomized list of sentences. The corpus is given below in (7).

Each sentence here is ambiguous in the sense that the initial adverb (or locative adjunct) can be taken to modify the verb of the relative clause directly following it (interpretation A: 'The scarf that I knitted last year was stolen.'), or it can modify the verb of the matrix clause (interpretation B: 'The scarf that I knitted was stolen last year.').

---

[3]Mineharu Nakayama has also suggested that the relatively higher peak height for Peak 2 in the RB structure may give listeners the impression of a pause. This is certainly a possibility, since is has been documented in English at least that adjacent peaks with about the same prominence can give the impression of a salient juncture or break.

(7) a. Temporal-*eri'maki* set:

kyo'nen a'nda eri'maki-ga nusuma'reta
last year knitted scarf-NOM was stolen

yuube a'nda eri'maki-ga nusuma'reta
last night knitted scarf-NOM was stolen

b. Locative-*eri'maki* set:

Me'jiro-de a'nda eri'maki-ga nusuma'reta
Mejiro-LOC knitted scarf-NOM was stolen

Ueno-de a'nda eri'maki-ga nusuma'reta
Ueno-LOC knitted scarf-NOM was stolen

The list of randomized sentences was presented to the speakers in normal Japanese orthography (kanji and kana), with a cue beneath indicating which meaning they should utter, as exemplified in (8). The speakers were asked to produce each sentence with the meaning indicated in the cue.

(8) yuube a'nda eri'maki-ga nusuma'reta.
(yuube a'nda) 'You knitted it last night.'

[actual cues were of course written in Japanese orthography without accompanying English translation.]

The speakers were given instructions for the task beforehand, presented in Japanese orthography, which outlined the ambiguity of the sentences and the contexts in which each interpretation might be uttered. Each speaker was allowed to practice before being recorded, but no verbal feedback or prompts were given. Ten tokens of each interpretation of all the utterances were elicited from each speaker, totaling 240 utterances (3 speakers x 2 branching x 2 type adjuncts x 2 word1 accentuations x 10 tokens). The tokens were analyzed as outlined in §3.1.1 above.

Results and discussion

The results from the two sets are essentially the same, as predicted from the fact that the accentuation of the components are identical, the only difference between the two is the type (temporal or locative) of adjunct being adjoined to S.[4] In light of this similarity, the discussion below will focus on the temporal-*eri'maki* set.

Figure 11 shows sample contours for both interpretations. The results for this experiment showed more consistency across the three speakers than did the results for the noun phrase case. As with the previous experiment, the behavior of the peak following accented versus unaccented first phrases was examined in order to determine whether or not there exists a downstep-blocking major phrase break between Peaks 1 and 2 in the right-branching structure. In contrast to the noun phrases, it may not be appropriate to characterize these relative clause constructions as 'left-branching' or 'right-branching', since relative clauses in Japanese are all left-branching. The distinction to be made between the two structures here is characterized by Uyeno et al. (1980) as 'left-branching', in which the initial adverb (or locative adjunct) modifies the verb of the relative clause, versus 'center-embedding', in which the adverb modifies the verb of the matrix clause. As far as

[4]The syntactic structures of these sentences are given in Figure 18 below.

pitch range relationships are concerned, these two structures behave very similarly to LB and RB noun phrases, respectively. In the following description, I will refer to the left-branching relative clause as interpretation A, and the center-embedding relative clause as interpretation B.

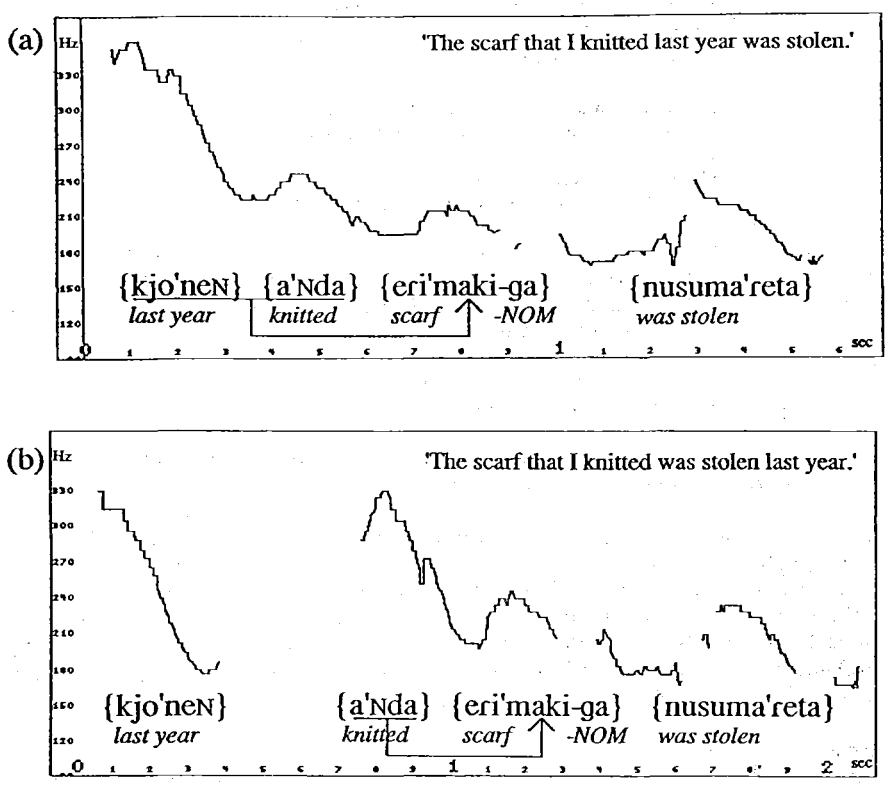


Figure 11. Typical fundamental frequency contours of interpretations A (top) and B (bottom) of the ambiguous string [kyo'nen a'nda eri'maki-ga nusuma'reta].

In order to examine the downstep relationship between Peak 1 (initial adverb or locative adjunct) and Peak 2 (verb of the relative clause), plots of the frequency of Peak 2 when following an accented and unaccented initial word are shown in Figure 12, for one speaker (KS). Again, as with the noun phrases, because of the dephrasing of the initial unaccented words in the left-branching (A) interpretations, only the relationships of the peaks in the B type (analogous to the RB structure) could be examined.

In both sets it is clear that the height of Peak 2 (*a'nda*) is not significantly lower when following an accented word as opposed to an unaccented word (adverb: $t = -2.28$, $p > 0.01$; locative: $t = -1.51$, $p > 0.01$). This result was found for the other two speakers as well, for both sets of utterances. Thus, according to the criteria for the detection of downstep set by Poser (1984) and confirmed by Pierrehumbert and Beckman (1988), Kubozono (1988), and others, there does not appear to be any

downstep occurring between these two peaks. Rather, its application is blocked, and the pitch range is reset on Peak 2 at the start of a new major phrase. These results support the account by Beckman and Pierrehumbert as well as the claims of Selkirk and Tateishi (1991) who propose a major phrase break for these types of structures. They form an apparent contradiction to Kubozono's claim that metrical **boost is applied to a downstepped peak to create the pitch range expansion in** complex structures such as these.

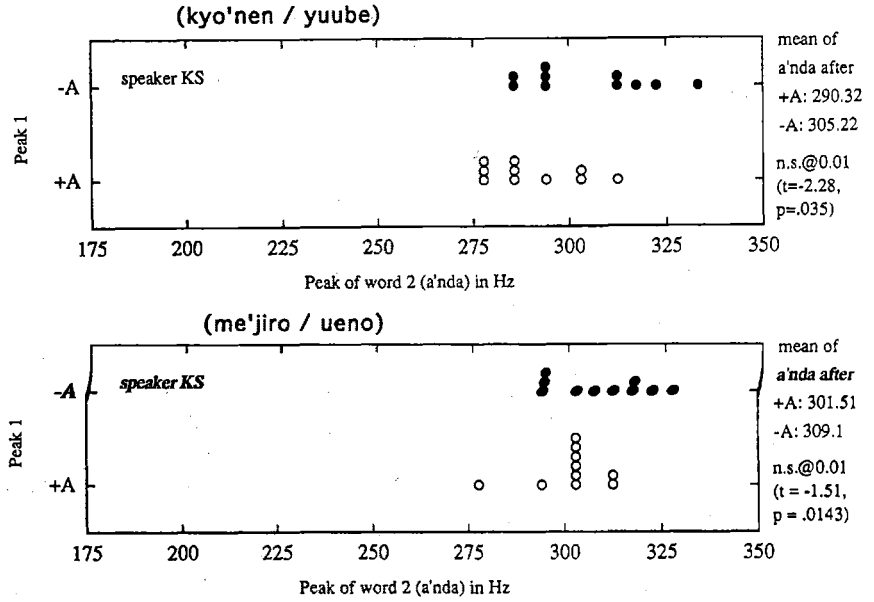


Figure 12. Fundamental frequency values of Peak 2 in interpretation B relative clause constructions [kyo'nen / yuube [a'nda] eri'maki-ga nusuma'reta] and [Me'jiro-de / Ueno-de [a'nda] eri'maki-ga nusuma'reta]. Filled circles indicate word 1 is -A, hollow circles indicate word 1 is +A. Speaker KS.

Since downstep is said to apply iteratively within the major phrase, it is useful to look at the relationship between the height of Peak 1 and Peak 3. If downstep is indeed blocked between Peaks 1 and 2, as the data seem to indicate, then one would expect that the accentedness of Peak 1 would also not have an effect on the height of Peak 3 (*eri'maki*). If, on the other hand, downstep does apply between Peaks 1 and 2, then it is expected that the difference in the height of Peak 3, even in the type B relative clauses, would show some signs of downstep chaining onto it. The result of a comparison of Peaks 1 and 3 is shown in Figure 13 for speaker AM, a representative case.

It is clear from Figure 13 that there is no effect of downstep between Peaks 1 and 2 which may be chaining onto Peak 3 in the B type structure (adverb: $t = .189$, $p > 0.01$; locative: $t = -1.93$, $p > 0.01$). This result holds true for the two other speakers as well: for all speakers, the height of Peak 3 was not lower when following an initial accented word as opposed to an initial unaccented word. This can be described in Beckman and Pierrehumbert's framework by saying that there

is an major phrase boundary which blocks downstep between Peaks 1 and 2 in these type B constructions.

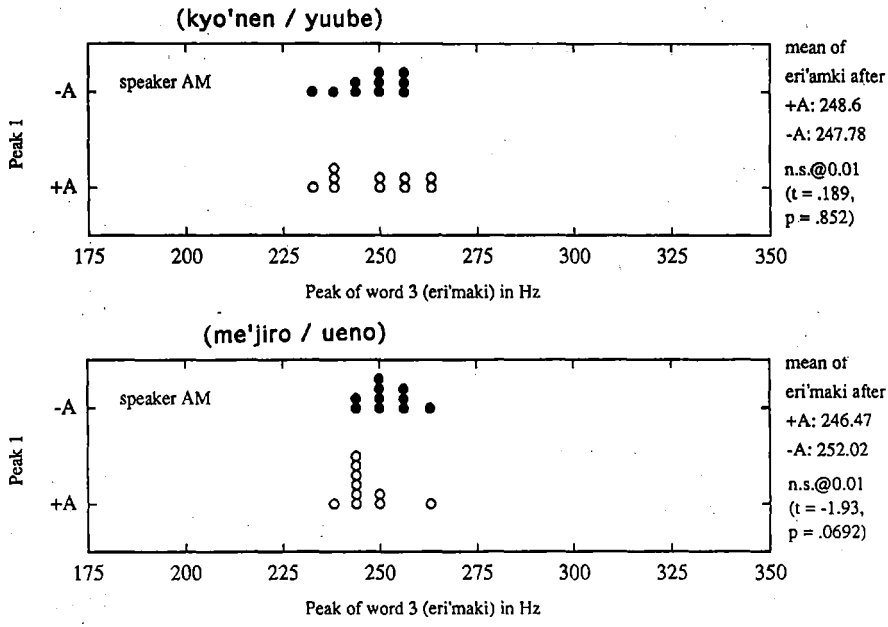


Figure 13. Fundamental frequency values of Peak 3 in interpretation B relative clause constructions [kyo'nen / yuube [a'nda] eri'maki-ga nusuma'reta] and [Me'jiro-de / Ueno-de [a'nda] eri'maki-ga nusuma'reta]. Filled circles indicate word 1 is -A, hollow circles indicate word 1 is +A. Speaker AM.

Another interesting result of this study concerns the scaling of initial peak height. Figure 14 shows frequency distributions of initial accented and unaccented peak heights for both branching structures. (Here, both the Temporal-*eri'maki* and Locative-*eri'maki* sets have been combined.) These distributions show that, for both accented and unaccented initial words, the height of the initial peak is significantly higher in type A constructions than it is in the type B constructions (-A: $t = 8.33$, $p < 0.01$; +A: $t = 8.12$, $p < 0.01$). These results hold true for the two other speakers as well, with the exception of the accented initial words for speaker KS, where the difference only approaches significance at the 0.01 level (MM-A: $t = 10.8$, $p < 0.01$; MM+A: $t = 6.63$, $p < 0.01$; KS-A: $t = 3.4$, $p < 0.01$; KS+A: $t = 2.62$, $p < 0.05$).

Such findings are interesting since they suggest that the speaker needs to look ahead at the syntactic configuration of an utterance even before s/he actually utters the first word. The overall 'mental plan' of the sentence will thus effect the pitch range scaling of even the leftmost component. If the initial adjunct modifies the immediately following verb, the speaker will utter it with a higher overall pitch than if it modifies the matrix verb three words later. This variability in initial peak scaling according to the syntactic structure of the rest of the utterance is similar to the findings of Ladd for English (Ladd, 1988; Ladd & Johnson, 1987). Ladd

states that "it is both necessary and appropriate to enrich the phonological representation of intonation in order to express the fact that syntactic organisation may be signalled intonationally in fine differences of peak height." (Ladd, 1986: 329) It was this notion which led him to propose his metrical representation of pitch range, in which the syntactic structure is encoded into an overall phonological 'plan' of the utterance, as was briefly outlined in §2.4 above. Ladd proposes that downstep can be modeled by a high-low branching node as in (9).

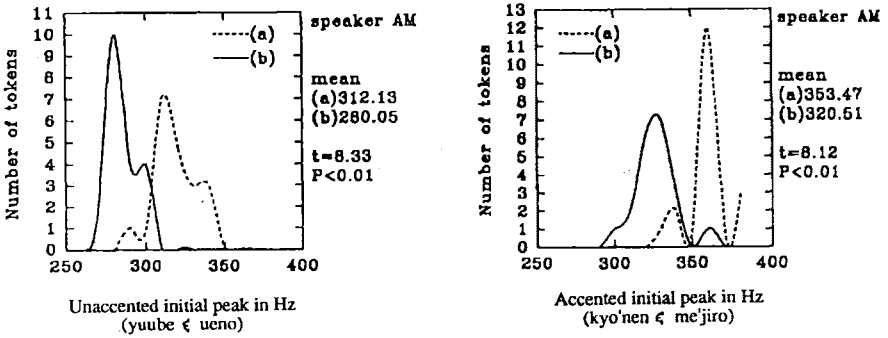
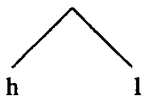


Figure 14. Frequency distributions of initial peak height in type A (dotted line) and type B (solid line) relative clause constructions. Temporal-*eri'maki* and Locative-*eri'maki* sets combined. Speaker AM.

(9)



The metrical tree of an utterance will mimic the syntactic structure by the direction of branching or the depth of embedding. The pitch register relationships which are in effect a consequence of this structure are represented by different configurations of the downstepping h-l sequence, or the non-downstepping l-h. Such a model is attractive since it by definition encodes pitch register relationships into its representation. In a model such as the one proposed by Beckman and Pierrehumbert, there is no way of predicting the relative heights of the first and second words in the two structures directly from the prosodic representation itself. One would have to rely on a discourse structure or some other syntactically sensitive structure to provide the information of relative pitch heights. Thus, Ladd's model does seem attractive, if one wants to encode the notions of syntactic branching into the phonological representation. However, his model is only one of relative pitch relationships, and until enough quantitative data are presented and modeled to determine exactly how to translate these 'h's and 'l's into actual pitch values, the proposal still remains speculative.

This experiment involving ambiguous scope of adverbial (or locative adjunct) modification in relative clause constructions was carried out in an attempt to examine Kubozono's conjectures about the multiple application of metrical boost in deeply embedded right-branching structures. Kubozono (1989, 1992) suggests

that the same notion of metrical boost seen applying to already downstepped peaks in right-branching noun phrases can be expanded to explain any boost in pitch range at the left edge of a right syntactic branch. He proposes that, with metrical boost as an n-ary process, it is possible "to eliminate the conventional rule of pitch register reset from the intonational system of Japanese." (1992: 382) While we saw evidence of metrical boost applying to a downstepped peak for some speakers with noun phrases, it is not clear that this is an appropriate characterization of the more complex utterances. Results from the second experiment showed that there was in fact no downstep occurring between Peaks 1 and 2 in the type B utterances for any of the speakers, nor did there seem to be any effects of downstep chaining onto Peak 3. These trends were found irrespective of the syntactic status (temporal or locative adjunct) of the first element. This suggests that the prosodic structure of these type B utterances is best described as consisting of two separate major phrases, at whose boundary the pitch range is reset, rather than boosted. Until more work is done on quantifying the phenomenon of metrical boost, and the effects of its multiple application, the account of Pierrehumbert and Beckman (1988) provides a clearer, more straightforward explanation of the data.

3.2.2 Perception

Rationale

As with the noun phrases, the perception portion of this experiment was conducted in order to confirm that the production differences between type A and type B structures are perceptually salient. Again, in accordance with the proposals of Uyeno et al. (1980) and Azuma & Tsukuma (1990), we would predict that a large positive difference between the height of the first two peaks is more likely to cue the A interpretation, while a negative, very small positive difference, or no difference will cue the B interpretation.

Methods

Eleven native listeners of Tokyo Japanese (or near Standard) participated in this experiment. One speaker was excluded since her results (4% correct) indicated that she could not attend to the task. Essentially the same method as the noun phrase perception experiment was used with these relative clause constructions. In this case, only the utterances in the Temporal-*eri'maki* set were used as stimuli, and every token was presented only once. This gave a total of 120 stimuli in all (3 speakers x 2 branchings x 2 word1 accentuations x 10 tokens). Each of the ten participants had judgments of 86% to 98% correct.

Results and discussion

As with the data of the relative clause construction production test, the results of this perception experiment were more consistent than the noun phrase perception. There was better agreement among listeners as to whether the token was the type A or type B interpretation, and there were no outlying tokens. Figure 15 shows listener judgment plotted as a function of the difference in the height of Peaks 1 and 2. Again, due to dephrasing, only the all accented type *kyo'nen a'nda eri'maki-ga nusuma'reta* could be examined.

It is clear from this graph that listeners were able to identify the structure of the utterance (interpretation A or B) with much greater success than they had with the noun phrases. Therefore, since most of the points are clustered around the two extremes, it is difficult to see if there is a transition from 'definitely A' (upper right) to 'definitely B' (lower left) as was clear with the noun phrases. There does seem to be a slight tendency for the A tokens with a smaller ERB value to extend toward

the middle of the graph, showing the traces of a correlation between the difference in height and the listener judgment. Figure 16 shows a similar clear-cut pattern in the relation between pause and the listener choice.

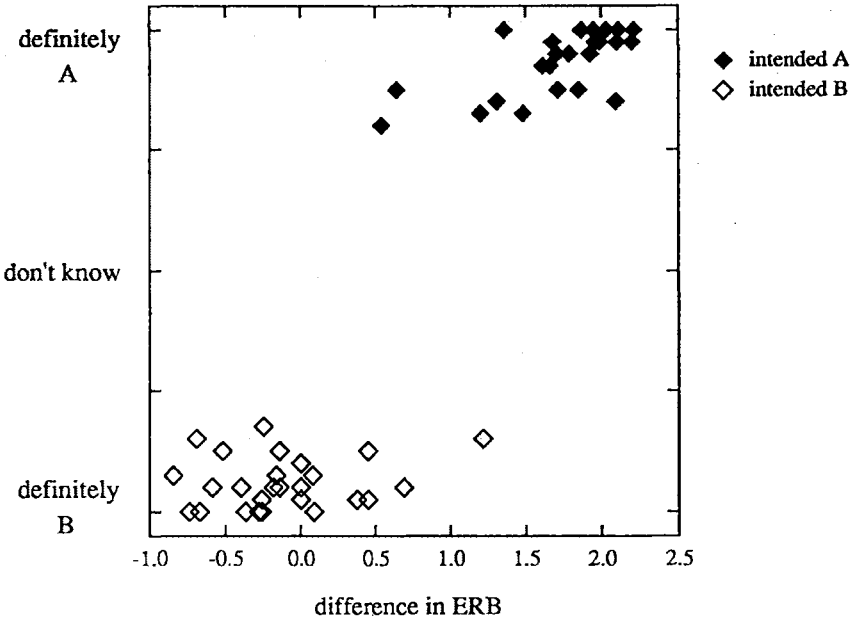


Figure 15. Listener choice from 'definitely A' to 'definitely B' plotted as a function of the difference in ERB between Peaks 1 and 2 for [kyo'nen a'nda eri'maki-ga nusuma're ta] relative clause constructions.

This graph clearly shows that none of the intended A tokens had a pause occurring between Peaks 1 and 2, while all of the intended B tokens contained a pause.[5] It might therefore seem that it is the pause only which distinguishes the two interpretations. However, the fact that the judgments for tokens in both A and B categories vary suggests that there is something in addition to this which influences native listeners' perception of these structures. A table outlining the cues which listeners after the experiment named as being relevant to disambiguation is shown in (10).

(10)

attended to pause (for B) only:	1 listener
attended to peak height difference only:	1 listener
attended to both:	6 listeners
neither:	2 listeners

[5]None of the speakers in this study produced type B utterances without a pause, however, it is possible to do so. Speaker YO in a follow up experiment uttered the same constructions with little or no pause, due to the rapid rate of speech. Also see Uyeno et al. (1980) and Azuma & Tsukuma (1990, 1991) for B type utterances without pauses.

In this perception test, listeners tended to be more aware of the difference in peak heights, and felt that, together with a pause, this could cue the appropriate interpretation. However, as noted before, these impressions may not be a reliable indicator of the actual cues which native speakers listen for.

It is therefore possible to conclude that, as with the noun phrases, the perceptual cue to the two interpretations here seems not to be a single factor, but rather a combination of cues such as the relative height of Peaks 1 and 2, the presence of a pause, as well as other prosodic or segmental variations not examined in this study such as duration or voice quality differences.

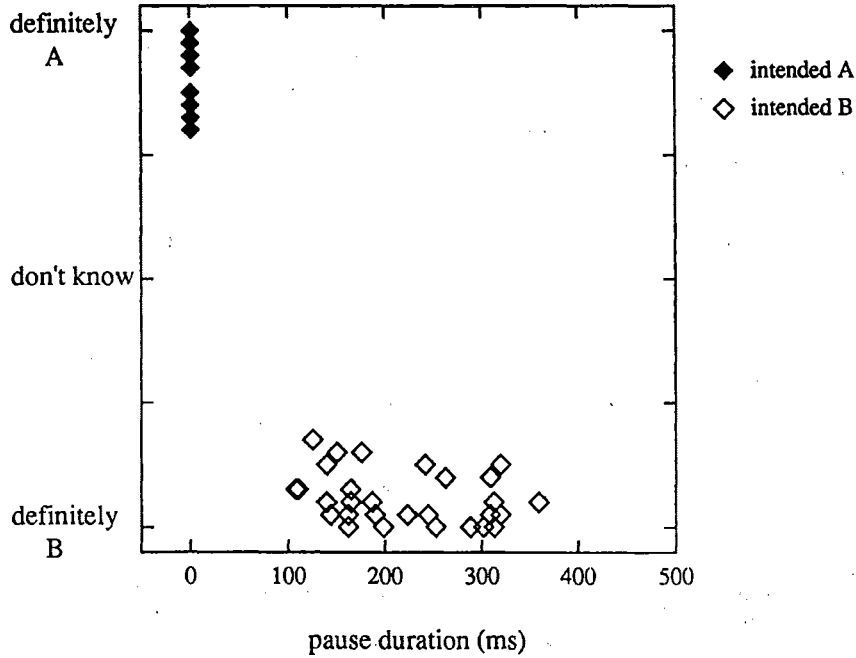


Figure 16. Listener choice from 'definitely A' to 'definitely B' plotted as a function of length of pause between Peaks 1 and 2.

4 Discussion

The present study examined downstep relationships between the first two peaks of ambiguous noun phrases and relative clause constructions. It was carried out in an attempt to replicate Kubozono's (1988) findings for noun phrases that downstep can apply across a syntactic boundary. This study was also designed to test his hypothesis for more complex constructions like relative clauses. The results show that, for noun phrases, two of the three Tokyo speakers behaved as did Kubozono's speaker, downstepping Peak 2 in relation to Peak 1 even in the right-branching structure. However, the third speaker did not show any evidence of downstep between the two peaks in a majority of her utterances, but rather had a major phrase break there which blocked downstep and induced a reset in pitch range. This suggests that the prosodic manifestations cannot be directly attributed

to a certain syntactic branching configuration, and that speakers may use different strategies involving variant prosodic phrasing or optional application of metrical boost. For relative clauses, on the other hand, all speakers used in this study behaved according to the Beckman and Pierrehumbert model. In these structures with a deep syntactic boundary, speakers chose to disambiguate the structures by means of the major phrasing. An examination of the initial peak scaling in these constructions suggests that the choice of the fundamental frequency value on this peak depends on the syntactic branching of the utterance. This resembles other findings by Ladd which originally motivated his metrical representation of pitch register. Perception tests on each of the corpora suggest that listeners use both the difference in the height of Peaks 1 and 2 as well as the presence of a pause as cues to disambiguate the structure of the continuous stream of segmental information.

5 Implications for syntax-prosody mapping

It is clear both from previous descriptions of structural ambiguity and from the results of the present study that the prosodic manifestation of an utterance is indeed influenced by its syntactic structure. While this has been taken as a matter of fact by most researchers, less attention has been paid to how exactly this mapping between the syntax and prosody is achieved. The following discussion will outline briefly some previous accounts and algorithms of this mapping, and examine whether they can account for the results of the present study.

One attempt at characterizing the relationship between the syntactic and prosodic structures has already been outlined above in some detail. In this account, proposed by Kubozono (1988, 1989, 1992), the prosody directly accesses certain syntactic configurations — here, the marked left-branching structure. He claims, as was noted above, that an upstep mechanism called 'metrical boost' applies to a minor phrase which finds itself as the leftmost constituent of a right-branch. Consider in Figure 17 the structures of noun phrases such as those discussed in this study (assuming an X' type of syntactic representation).

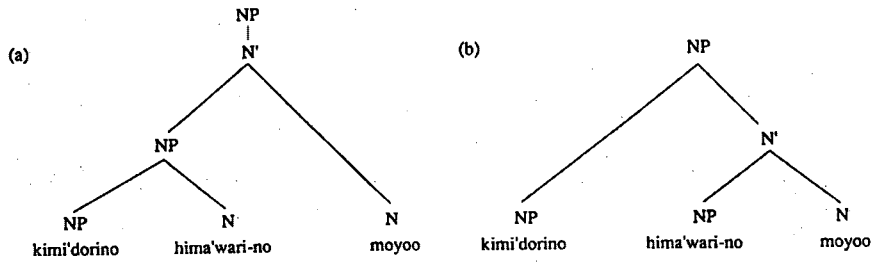


Figure 17. Syntactic structures of left-branching (a) and right-branching (b) noun phrases.

Assuming these structures, the account proposed by Kubozono is straightforward — the minor phrase which falls on the right-branch in (b) will receive a metrical boost to raise its already downstepped peak (cf. Figure 4b). Kubozono chooses to encode the syntactic structure into a recursive prosodic representation which then triggers metrical boost. However, it is also possible to access the syntax directly without encoding it in the phonology first. The end result will be the same.

The question then becomes whether such an account will work for constructions of greater complexity than noun phrases and with deeper levels of embedding. Figure 18 gives the structures of the relative clause constructions used in this study.

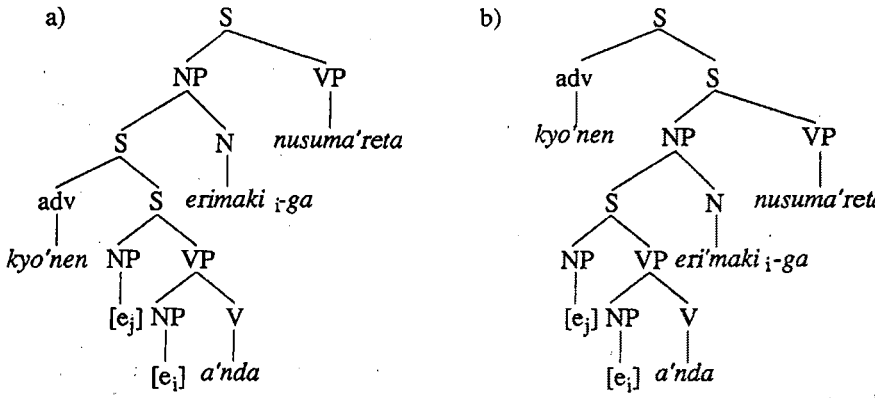


Figure 18. Syntactic structures of type A (a) and type B (b) relative clause constructions.

Kubozono (1992) offers the suggestion that the mechanism of metrical boost is an n-ary process which may apply multiply to right-branching configurations. The structures in Figure 18 are restated in (11) as bracketed strings for easier recognition of the right-branches.

(11) a. [[[kyo'nen [e_j [e_i a'nda]]] eri'maki$_i$-ga] nusuma'reta]
'The scarf that I knitted last year was stolen.'

b. [kyo'nen [[[e_j [e_i a'nda]] eri'maki$_i$-ga] nusuma'reta]]
'The scarf that I knitted was stolen last year.'

It is quite apparent that the type of syntax one assumes will effect the location and depth of the right-branching elements. Even in these fairly basic representations of the two constructions, given a framework which assumes various empty categories, there will be many right-branching nodes that a theory such as Kubozono's is probably not prepared to deal with. If we simplify the representation further, as shown in (12), we have a structure which is closer to that which Kubozono had in mind.[6]

(12) a. [[[kyo'nen a'nda] eri'maki-ga] nusuma'reta]
b. [kyo'nen [[[a'nda] eri'maki-ga] nusuma'reta]]

Applied to such a structure, metrical boost would apply three times on the downstepped verb *a'nda* in (b), causing the peak to be boosted quite a bit. The results discussed above indicate that there is no sign of downstep occurring between Peaks 1 and 2, but rather support an interpretation that the pitch range is being reset completely. While the proposal of a multiple application of metrical

[6]The reader is referred to a similar example in (8b) Kubozono (1992).

boost which in effect 'undoes' downstep totally if applied enough times is still plausible, without more details of the exact quantitative nature of this mechanism and its cumulative effects, it is difficult to support or reject this account.

Other authors have proposed alternative methods of mapping the syntax onto the prosody. Most noteworthy is the edge-based theory proposed by Selkirk (1986) and argued for Japanese by Selkirk and Tateishi (1988, 1991). According to their proposal, the boundary of a major phrase — at which pitch range is reset — corresponds to the left-edge of a maximal projection in the syntactic hierarchy.[7] In such a model, the prosodic structure is influenced by the syntax, but not directly. Selkirk and Tateishi (1988, 1991) have argued that this account can describe the syntactic structures which they examined in their study. It is an empirical question then whether their account will correctly predict the phrasings found in the present study. Given a representation such as that in Figure 17, we would predict that the string would consist of a single major phrase in the LB structure, but two separate major phrases in the RB structure: (kimi'dorino)(hima'wari-no moyoo). This prediction is confirmed by the results of only one speaker in the present study. It fails to explain why we observe an increased prominence on the already downstepped Peak 2 in right-branching structures for the two speakers, who produced the RB strings with just one major phrase. Therefore, in the face of such inter- and intra-speaker variability, such an edge-based mapping theory which relies solely on the syntactic structure will run into trouble.

Still more difficulty for their account arises when we examine the structures of the relative clause constructions (cf. Figure 18). Without going into great detail about how their mapping algorithm might apply to such structures, the crucial thing to note is that, at the level of the most subordinate S, the structures of the two interpretations are identical. Thus, it follows that whatever might apply or not apply to one structure must hold for the other, rendering them virtually identical for such mapping algorithms. Considering only this S, if we assume that phonologically empty nodes are place holders and can serve as the left-edge of a X^{max}, then the verb *a'nda* would start a new major phrase in both cases. Likewise, if we choose to ignore empty nodes, then there is nothing on the left-edge of a X^{max} in either structure which would predict a pitch range reset.[8] Therefore, the problem that such relative clause constructions with ambiguous scope of adjunct modification hold for such edge-based mapping theories, in virtually any syntactic framework, is that what the algorithm predicts for one structure will be identical to that which it predicts for the other structure.

The last approach to syntax-prosody mapping which I will address here is that of Nespor and Vogel (1986). Theirs is a relation-based theory which holds the notions of head and complement crucial to the relation between syntactic and prosodic structures. They define the phonological phrase and the intonational phrase as the following:

(13) Phonological phrase domain:
> Consists of a clitic group which contains a lexical head (X) and all clitic groups on its nonrecursive side up to the clitic group which contains another head outside of X^{max}. (paraphrased, Nespor & Vogel, 1986:168)

[7]The edge parameter is set for 'left' in Japanese.

[8]Selkirk and Shen (1990) chose to ignore empty categories, claiming that a phonologically null trace has no effect on the syntax-prosody mapping.

Intonational phrase domain:

> An intonational phrase may consist of all the phonological phrases in a string that is not structurally attached to the sentence tree at the level of s-structure, or any remaining sequence of adjacent phonological phrases in a root sentence. (Nespor & Vogel, 1986:189)

First let us consider the relative clause constructions shown in Figure 18. If we take the phonological phrase to be equivalent to the accentual phrase (Vogel, personal communication), and the intonational phrase to be the major phrase, the correct phrasing is predicted. Each of the words in both structures (a) and (b) is a lexical head, and it happens that they are separated from one another by a maximal projection. Therefore, the prediction that each word forms a single accentual (phonological) phrase is correct. However, note that each of these words are accented and thus tend to form their own phrase (cf. §2.2). If one of the words had been unaccented, as in *yuube a'nda eri'maki-ga nusuma'reta*, the prediction would be incorrect since *yuube* is dephrased together with the following verb. Therefore, while the algorithm for determining accentual phrases holds in this structure for accented words, substitution of unaccented words will complicate matters.

Regarding the intonational (major) phrase, the predictions of Nespor and Vogel's mapping algorithm do predict the correct distinction between the two structures in Figure 18. This theory predicts that structure (a) will form one major phrase since all of the accentual (phonological) phrases are part of the root sentence at s-structure, whereas structure (b) would form two major phrases since the adverb is not attached to the root sentence. It is not clear to me if the fact that both the adverb and its sister S are attached to a higher S has any bearing on the validity of their prediction.

However, if we attempt to apply Nespor and Vogel's mapping algorithm to the noun phrase constructions shown in Figure 17, we immediately run into trouble. It cannot account for the speaker variability discussed above, and also has problems with describing dephrasing of unaccented words in the string.

It is obvious from the discussion above that the results of this study present problems for all of the major theories of syntax-prosody mapping: evidence of the fact that the influence syntax has on prosodic structure is anything but straightforward. While Selkirk and Tateishi and Nespor and Vogel's mapping algorithms need to undergo major revision to account for the present data, Kubozono's proposal may be the direction in which we should look when the phenomenon of metrical boost has been more carefully examined and documented.

6 Conclusion

Syntactic structure can indeed influence the prosodic realization of an utterance. This has been shown not only in the present experiments but in numerous previous studies of several languages. The more relevant issues are exactly *how* it does this and to what extent we should encode this influence into our phonological representation. The present study looked at structurally ambiguous utterances in Japanese and how disambiguation of these is achieved via the fundamental frequency contour. Structures involving left-branching nodes are characterized by a downstepping of peaks resembling a staircase pattern. However, the right-branching structures examined here were not uniform in how they were realized prosodically. In those structures in which the right-branch is not deeply embedded, as with noun phrases, two of three speakers produced one major phrase with an

additional boost on Peak 2. However, in the structures with the deeply embedded right branch, as in the relative clause constructions, all speakers produced two major phrases accompanied by a reset in pitch range. Given this apparent lack of consistency among the two constructions, in an account of the relationship between syntactic and prosodic structures, the depth of embedding is an important issue to consider. Another thing to consider is how we want to represent the influence of the syntax on prosodic structure. The theory of prosodic structure advocated by Ladd and Kubozono aims to encode syntactically sensitive variations in pitch register into the phonological representation, while alternative theories such as that of Beckman and Pierrehumbert wish to leave those decisions up to the pragmatics or discourse structure. A model that encodes the syntax necessarily complicates the phonological representation and cannot account for speaker variation. On the other hand, a model which leaves everything up to the pragmatics, without a concrete model of pragmatic or discourse structure, leaves too many degrees of freedom, and lessens the predictive capabilities.

It is clear that there is much room for future research in this area. An elaboration and more careful documentation of the phenomenon of metrical boost is necessary in order to examine its relation with syntactic branching structures, or more likely, pragmatic or discourse structures which are probably only vaguely related to syntactic constituency. Also, downstepping patterns in utterances with more diverse syntactic structures need to be thoroughly examined in order to assess the contribution of depth of embedding to the prosodic juncture between adjacent words. Lastly, a coherent model of discourse structure and its relation to the prosodic structure would prove extremely useful in determining the relative weights of contribution from the syntax and pragmatics to the overall prosodic realization of an utterance.

References

Azuma, J. & Tsukuma, Y. (1990) Prosodic features marking the major syntactic boundary of Japanese: A study on syntactically ambiguous sentences of the Kinki dialect. *Proceedings of the International Conference on Spoken Language Processing*, Kobe, 453-455.

Azuma, J. & Tsukuma, Y. (1991) Role of F0 and pause in disambiguating syntactically ambiguous Japanese sentences. *Proceedings of the XIIth International Congress of Phonetic Sciences*, 3: 274-277.

Beckman, M.E. & Pierrehumbert, J.B. (1986) Intonational structure in Japanese and English. *Phonology Yearbook*, **3**, 255-309.

Cooper, W.E., Paccia, J.M. & Lapointe, S.G. (1978) Hierarchical Coding in Speech Timing. *Cognitive Psychology*, **10**, 154-177.

Hermes, D.J. & van Gestel, J.C. (1991) The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, **90**, 97-102.

Jun, S.-A. (1993) *The phonetics and phonology of Korean prosody*. Ph.D. dissertation (Ohio State University).

Klatt, D.H. (1975) Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, 129-140.

Kori, S. (1992) Nihongo bun'onchoo no kenkyuu kadai. *Proceedings of the International Symposium on Prosody*. Nara, Japan.

Kubozono, H. (1988) *The organization of Japanese prosody*. Ph.D. dissertation (University of Edinburgh).

Kubozono, H. (1989) Syntactic and rhythmic effects on downstep in Japanese. *Phonology*, **6**, 39-67.

Kubozono, H. (1992) Modeling syntactic effects on downstep in Japanese. In *Papers in Laboratory Phonology II: Segment, Gesture, Tone* (G.J. Docherty & D.R. Ladd, editors), pp. 368-387. Cambridge: Cambridge University Press.

Ladd, D.R. (1986) Intonational phrasing: The case for recursive prosodic structure. *Phonology Yearbook*, **3**, 311-340.

Ladd, D.R. (1988) Declination "reset" and the hierarchical organization of utterances. *Journal of the Acoustical Society of America*, **84**, 530-544.

Ladd, D.R. (1990) Metrical representation of pitch register. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (J. Kingston & M.E. Beckman, editors), pp. 35-57. Cambridge: Cambridge University Press.

Ladd, D.R. (1993) In defense of a metrical theory of downstep. In *The Phonology of Tone: The Representation of Tonal Register* (H. van der Hulst & K. Snider, editors), pp. 109-132. Mouton de Gruyter.

Ladd, D.R. & Johnson, C. (1987) 'Metrical' factors in the scaling of sentence-initial accent peaks. *Phonetica*, **44**, 238-245.

Lehiste, I. (1973) Phonetic disambiguation of syntactic ambiguity. *Glossa*, **7**, 107-122.

Lehiste, I., Olive, J.P. & Streeter, L.A. (1976) The role of duration in disambiguating syntactically ambiguous utterances. *Journal of the Acoustical Society of America*, **60**, 1199-1202.

Maekawa, K. (1991) Perception of intonational characteristics of WH and non-WH question in Tokyo Japanese. *Proceedings of the XIIth International Congress of Phonetic Sciences*, 4/5: 202-205.

McCawley, J. (1968) *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.

Moore, B.C.J. & Glasberg, B.R. (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, **74**, 750-753.

Nespor, M. and I. Vogel (1986) *Prosodic Phonology*. Dordrecht: Foris Publications.

Pierrehumbert, J.B. & Beckman, M.E. (1988) *Japanese Tone Structure*. Cambridge, Massachusetts: MIT Press.

Poser, W. (1984) *The Phonetics and Phonology of Tone and Intonation in Japanese*. Ph.D. dissertation (Massachusetts Institute of Technology).

Poser, W. (1990) Word-internal phrase boundary in Japanese. In *The Phonology-Syntax Connection* (S. Inkelas & D. Zec, editors), pp. 279-287. Chicago: University of Chicago Press.

Selkirk, E. (1984) *Phonology and Syntax: The Relation between Sounds and Structure*. Cambridge, Massachusetts: MIT Press.

Selkirk, E. (1986) On Derived Domains in Sentence Phonology. *Phonology*, **3**, 371-405.

Selkirk, E. & Tateishi, K. (1988) Constraints on minor phrase formation in Japanese. *Proceedings of the Chicago Linguistic Society*, **24**, 316-336.

Selkirk, E. & Tateishi, K. (1991) Syntax and downstep in Japanese. In *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda* (C. Georgopoulous & R. Ishihara, editors), pp. 519-543. Dordrecht: Kluwer Academic Publishers.

Streeter, L.A. (1978) Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*, **64**, 1582-1592.

Terken, J. & Collier, R. (1992) Syntactic influences on prosody. In *Speech Perception, Production and Linguistic Structure* (Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka, editors), pp. 427-438. Tokyo: IOS Press.

Uyeno, T., Hayashibe, H. & Imai, K. (1979) On pitch contours of declarative, complex sentences in Japanese. *Research Institute of Logopedics and Phoniatrics*, 13, 175-187. University of Tokyo.

Uyeno, T., Hayashibe, H., Imai, K., Imagawa, H. & Kiritani, S. (1980) Comprehension of relative clause construction and pitch contours in Japanese. *Research Institute of Logopedics and Phoniatrics*, 14, 225-236. University of Tokyo.

Uyeno, T., Hayashibe, H., Imai, K., Imagawa, H. & Kiritani, S. (1981) Syntactic structures and prosody in Japanese: A study on pitch contours and the pauses at phrase boundaries. *Research Institute of Logopedics and Phoniatrics*, 15, 91-108. University of Tokyo.

Venditti, J.J. & Yamashita, H. (in preparation) The prosodic characteristics of temporarily ambiguous constructions in Japanese. ms. Ohio State University.